DOCUMENT RESUME

ED 252 230 IR 050 970

AUTHOR	David, Martin; And Others
TITLE	Archival Preservation of Machine-Readable Records: The Final Report of the Wisconsin Survey of Machine-Readable Public Records.
INSTITUTION	Wisconsin State Historical Society, Madison.
SPONS AGENCY	National Historical Publications and Records Commission, Washington, DC.
REPORT NO	ISBN-0-87020-212-X
PUB DATE	81
GRANT	NHPRC-80-8
NOTE	54p.; A pilot program to accession machine-readable public records of Wisconsin State Agencies. A cooperative project between the Archives Division of the State Historical Society of Wisconsin and the Data and Program Library Service of the University of Wisconsin-Madison, 1979-1981. For related documents, see IR 050 970-973.
AVAILABLE FROM	State Historical Society of Wisconsin, 816 State Street, Madison, WI 53706 ($5.00 per copy).
PUB TYPE	Reports - Evaluative/Feasibility (142)
EDRS PRICE	MF01/PC03 Plus Postage.
DESCRIPTORS	*Archives; Computers; Data Processing; *Government Publications; Information Retrieval; *Information Storage; Records (Forms); *State Agencies; State Programs; State Surveys; Statewide Planning
IDENTIFIERS	*Machine Readable Bibliographic Data Bases; Machine Readable Cataloging; Records Management; *Wisconsin

ABSTRACT

Between November 1, 1979, and April 30, 1981, the Wisconsin Survey of Machine-Readable Public Records was carried out to identify the technical, intellectual, and administrative problems associated with a records management and archival program for machine-readable records. The survey identified machine-readable records in several state agencies; evaluated existing records management, disposition, and retention policies governing machine-readable records; and developed a set of recommendations for improving records management and archival control of these materials. This final report summarizes the survey project. In Part One the history of the project and the strategies employed to inventory, appraise, and accession machine-readable records in the state of Wisconsin are described. Part Two describes the findings of the records survey. Part Three contains recommendations for state agency administrators, legislators, and archivists for establishing a machine-readable records program for state archives, dealing with pre-archival control, legislative support, documentation standards, data base management systems design, and administrative expertise. (THC)

ED252230

Archival Preservation of Machine Readable Records: A Final Report

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Max J. Evans

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

STATE HISTORICAL SOCIETY OF WISCONSIN

IR050970

ARCHIVAL PRESERVATION OF MACHINE-READABLE RECORDS:

THE FINAL REPORT OF THE WISCONSIN SURVEY OF MACHINE-READABLE PUBLIC RECORDS

A Pilot Program to Accession Machine-Readable
Public Records of Wisconsin State Agencies
A Cooperative Project Between the Archives Division
of the State Historical Society of Wisconsin
and the Data and Program Library Service of the
University of Wisconsin-Madison, 1979-1981

Funded in Part by a Grant From the
National Historical Publications and Records Commission
Grant Number 80-8

MARTIN DAVID and F. GERALD HAM
Co-Principal Investigators

MAX J. EVANS and ALICE ROBBIN
Project Co-Directors

MARGARET L. HEDSTROM
Project Archivist

The State Historical Society of Wisconsin
Madison, 1981

Copyright 1981 by
THE STATE HISTORICAL SOCIETY OF WISCONSIN
All rights reserved.

Library of Congress Cataloging in Publication Data

Main Entry under title:

Archival preservation of machine-readable records.

"A cooperative project between the Archives Division of the State Historical Society of Wisconsin and the Data and Program Library Service of the University of Wisconsin-Madison, 1979-1981."

1. Archives--Wisconsin--Data processing. 2. Wisconsin--History--Sources. 3. Documents in machine-readable form. I. David, Martin Heidenhain. II. Ham, F. Gerald, 1930- . III. State Historical Society of Wisconsin. Division of Archives and Manuscripts. IV. University of Wisconsin--Madison. Data and Program Library Service.

CD3591.A72 027.5775 81-16721

ISBN 0-87020-212-X AACR2

PREFACE

The historical role of archivists, as the secondary custodians of records, has been to preserve all forms of documentation regardless of its recording medium, be it etched stone, cuneiform tablets, papyrus, animal skins, or paper. The development of inexpensive, mass-produced, shortlived paper, for example, has required archivists to develop techniques of assuring that historically valuable material will not be lost.

While each of these recording media created its own problems and archivists have responded with unique solutions, all of these records had this in common: the information they recorded consisted of visual symbols, most commonly ink on paper. These eye-readable documents are, in an increasingly computerized society, giving way to a method of recording information which is outside the experience of most of the current generation of archivists. Information is symbolized not by marks on paper, but by combinations of on-off signals which exist as electronic, magnetic, or light impulses on a variety of media, and which can be "read" only by machines. While computer enthusiasts and futurists talk of a "paperless" society, it is not clear when, if ever, society will reach a point when paper documents will not be used. It is clear, however, that many business and bureaucratic functions are being automated and that this trend will continue. The records created in the course of carrying out these functions will become increasingly available in machine-readable (MR) form.

Such is the background for the Wisconsin Survey of Machine-Readable Public Records. To meet the demands of a complex society, new information recording and storage technologies have been created. A new medium does not necessarily suggest a new methodology, however. Indeed, fundamental to this project is the assumption that public records, regardless of the recording medium, share certain legal and administrative characteristics which require that their management be governed by sound principles derived from past practices. Machine-readable records (MRR), like any records, must be managed and controlled while still in active use; their disposition must be scheduled to assure that information is not maintained unnecessarily and that currently necessary and historically valuable information is not lost; selected records must be preserved in archives; and archival records must be made available for research.

This discussion of the common features of records does not imply that there are no differences between MRR and more traditional paper and microform records. Indeed, it was the purpose of the project to identify the technical, intellectual, and administrative problems associated with a records management and archival program for MRR. Many of these problems are outside the scope of most archivists' experiences. Archivists, although often intimidated by talk of bits, bytes, bauds, and bugs, nevertheless have an obligation to develop sufficient expertise to meet the challenges MRR present.

Another assumption was that, although MRR bring with them a new set of problems, they also present some opportunities: information in MR form is recorded very densely, thus offering a potential for saving space; MRR are easily manipulated and analyzed, thus providing improved access for researchers; and MRR make "masking" individual data elements possible, thus ensuring confidentiality.

This report is an analysis of both the problems and opportunities presented by MRR based on the experiences in one state. It is one step toward a better understanding of the wealth of non-traditional documentation that must be made available to future generations of historical researchers.

The report represents a close collaboration between two institutions--the State Historical Society of Wisconsin and the Data and Program Library Service (DPLS) of the University of Wisconsin-Madison. The Historical Society has the legal responsibility for the records of the State of Wisconsin and many years of experience in appraising, processing, and providing access to public records; the DPLS has fifteen years' experience in developing documentation standards, bibliographic control, and machine-independent management and maintenance systems. The cooperation of these two institutions has made it possible to chart a course for the administration of MRR archives.

July 1981

ACKNOWLEDGEMENTS

This project owes a great deal to many individuals and organizations who cooperated fully with the project staff, helpfully supplying information and insights and opening doors to additional contacts. Thanks go to the data processing managers and records management personnel in the Wisconsin Department of Administration who helped define the scope and nature of data processing in state government, and to the records officers, data processing personnel, program officials, and many others in the departments of Public Instruction, Revenue, and Health and Social Services, for their assistance and cooperation.

We especially acknowledge the substantial contributions of Katherine Unertl who, working closely with the project staff, took sole responsibility for those aspects of the project which required direct access to the computer. She did her job with precision, devotion, and adeptness at problem-solving. The project staff also acknowledges all those who assisted with the workshops: those who helped prepare the handouts and assisted with the arrangements, and the speakers--James McDermott, Mary Ann Woodke, and Larry Travis, and especially Bruce Ambacher of the National Archives and Records Service Machine-Readable Records Division who substituted in an admirable fashion at the last moment for the keynote speaker. We are also grateful to those who assisted with the research, typing, editing, and production of this final report and the technical reports: Janie Cohen of the Data and Program Library Services staff and Karen Baumann, Karen Fitch, Paul Hass, and George Talbot of the Historical Society staff.

Finally, we express appreciation to the National Historical Publications and Records Commission for providing funding for this project and to the Commission staff for their encouragement and guidance.

TABLE OF CONTENTS

		Page
PREFACE		iii
ACKNOWLEDGEMENTS		v
EXECUTIVE SUMMARY		1
INTRODUCTION		5
PART ONE: HISTORY OF THE PROJECT		
1.	Preparation	9
2.	Survey and Analysis	9
2.1	Selecting the Sample	
2.2	Identifying Data Processing Systems and Machine-Readable Records	
2.3	Identifying Information about Files	
2.4	Describing the Data Systems and Machine-Readable Records	
2.5	Selecting Data Files	
3.	Accessioning	13
4.	Workshops	13
5.	Reporting and Other Activities	14
PART TWO: FINDINGS		
6.	Quantity and Nature of Machine-Readable Records	17
7.	Problem Areas and Implications	19
7.1	Identification of Systems and Master Files	
7.2	Revisions and Modifications of Automated Data Systems	
7.3	Maintenance and Preservation	
7.4	Documentation	
7.5	Retention	
7.6	Scheduling	
7.7	Appraisal	
7.8	Hardware and Software	
7.9	Accessioning and Processing	
7.10	Access	
7.11	Confidentiality	
7.12	State-Federal Data Transfer	
8.	General Issues	35
8.1	Attitudes	
8.2	Staff Expertise	
8.3	Costs	
PART THREE: AN ARCHIVAL PROGRAM FOR MACHINE-READABLE RECORDS		
9.	Requisites for a Machine-Readable Records Program	41
9.1	A Records Management Program	
9.2	Archival Capabilities to Handle Machine-Readable Records	
9.3	Access to Outside Technical Skills and Resources	
9.4	Legislative and Administrative Guidelines	
10.	Elements of a Machine-Readable Records Program	43
10.1	Pre-archival control	
10.2	Archival Review	
10.3	Archival Preservation and Management of Machine-Readable Records	
10.4	Access and Use	
11.	Concluding Remarks	49

EXECUTIVE SUMMARY

Between November 1, 1979, and April 30, 1981, the Archives Division of the State Historical Society of Wisconsin and the Data and Program Library Service of the University of Wisconsin-Madison carried out a cooperative project, the Wisconsin Survey of Machine-Readable Public Records. This Final Report documents the history of the project, reports its findings, and presents recommendations for establishing a state archival program to administer machine-readable records (MRR).

The following recommendations are made to state agency administrators, legislators, and other state archivists who intend to implement a MRR program:

(1) Pre-archival control over state agency MRR must be established without delay. State agencies must develop a records management program for MRR that is responsive to the needs of state agencies and at the same time assures the preservation of important records for future use. MRR must be inventoried and scheduled by the state agencies for future disposition. The fragile and ephemeral nature of MRR requires that an agency-instituted program of maintenance and preservation be immediately put into place. Agency records custodians must be trained in these new types of records and new technologies. The state archives can assist in all these activities. Policies, systematic procedures, and cost-sharing formulas for the transfer and preservation of MRR must be developed.

(2) There must be legislative recognition of the value of MR administrative records for secondary analysis. Statutes and administrative rules must provide a means to make the records accessible for scholarly research while safeguarding the rights of privacy of individuals whose activities are documented in the records. Though relatively few records are restricted, such records constitute a valuable resource and need to be exploited. However, the lack of consistent policies and of systematic procedures to permit the use of these records for statistical and other research activities creates a significant barrier to their utilization. Archivists must become involved in drafting legislation that would permit scholarly access to this valuable resource.

(3) Lack of documentation about the creation and processing of MRR deters access by the record-creating agency, the archival agency, and researchers. Standards for documenting MRR have been developed and are being instituted by agencies of the federal government. State agencies and archives should utilize these standards to improve the quality of documentation of their MRR.

(4) Means must be developed to extract selected data from the large and complex data base management systems being installed by state agencies. Unless these systems are designed to maintain historical or non-current information, much of the record of the agencies' activities will be lost. The social research community, the archives, and the state agencies will all benefit from cooperative efforts to design data base management systems which have the capability to create historical files for research at a low cost.

(5) State archives must develop a capacity to administer a program for preserving MRR. The archives must develop the in-house capabilities to accession, maintain, and provide access to MRR. Without the capability to respond to new records structures, storage media, and technologies, important public records will be irretrievably lost. Further, the future will see fewer public records created in traditional formats and media. The archival profession must respond to the challenge of the new technologies. Because MRR are often complex, dynamic, and tied to a rapidly changing technology, archivists must seek guidance and expertise. Expertise is available through organizations such as university-based data archives and the National Archives and Records Service (NARS). This expertise should be applied to administering MR public records produced by state government.

During the project a number of technical reports were issued. They are included as appendices to the copy of the Final Report submitted to the National Historical Publications and Records Commission. Each state archives has been sent a copy of the Final Report, but not the appendices. They are available at cost from the State Historical Society of Wisconsin and are recommended for their detailed description of the impact of automation on state agency records-keeping practices.

Appendix A. Technical Report #1. A Report on Data Processing and Machine-Readable Records in the Department of Public Instruction. (61 pages)

Appendix B. Technical Report #2. A Report on Data Processing and Machine-Readable Records in the Wisconsin Department of Revenue. (64 pages)

Appendix C. Technical Report #3. A Report on Data Processing and Machine-Readable Records in the Department of Health and Social Services. (89 pages)

Appendix D. Data Collection Form for Describing the Data System and Machine-Readable Records. (3 pages)

Appendix E. Understanding Automated Record Systems. Workbook #1 (16 pages) Techniques to Inventory, Describe, Appraise and Schedule Machine-Readable Records. Workbook #2 (15 pages)

Appendix F. Technical Report #4. The Social Utility of Personal Information. An Examination of and Recommendations for Statutory Protection and State Agency Policies and Practices Regarding Research Access to Confidential Records. (207 pages)

Appendix G. A User's Guide to the Machine-Readable Data File. (sample documentation) (35 pages)

The project also accessioned and processed five data files. During the description process, finding aids were created for each data file. The principal finding aids--the user's guides--and the data files are available at cost from the State Historical Society and the Data and Program Library Service. These files include:

1. Department of Public Instruction. Division of Management, Planning and Federal Services. Public School Enrollment, 1974-1975. [machine-readable data file]

2. Department of Public Instruction. Division of Management, Planning and Federal Services. Non-Public School Enrollment, 1974-1975. [machine-readable data file]

3. Department of Public Instruction. Division of Management, Planning and Federal Services. Ethnic Data, 1974-1975. [machine-readable data file]

4. Department of Public Instruction. Division of Management, Planning and Federal Services. Teacher Name and Employment File, 1970-1971. [machine-readable data file]

5. Department of Public Instruction. Division for Handicapped Children. Needs Assessment Survey, 1978-1979. [machine-readable data file]

INTRODUCTION

During the last three decades, computer technology has altered the records-keeping practices of local, state, and federal government agencies. Computer-assisted administrative, research, policy, and evaluation activities create a new form of public record, the machine-readable record (MRR). MRR offer potential solutions to a number of pressing archival problems, including the increased volume of paper records, difficulties of information retrieval and manipulation, privacy and confidentiality concerns, reproduction and dissemination, and linkage with other records series.

There are few archival programs to administer public records in machine-readable (MR) form (the National Archives of the United States and the Public Archives of Canada are notable exceptions). Therefore information on the quantity of MRR generated by state agencies, their contents, and how to gain access to them has been lacking, and archivists have yet to develop the knowledge, technical skill, and resources to preserve and disseminate these records.

This lack of information and expertise led the Archives Division of the State Historical Society of Wisconsin and the Data and Program Library Service of the University of Wisconsin-Madison to request funding from the National Historical Publications and Records Commission (NHPRC) to carry out a pilot project to inventory, appraise, and accession MRR from selected Wisconsin state agencies. The project brought together the expertise of public records archivists and specialists in MRR. The objectives were to develop an archival program for appraising, accessioning, preserving, and using MRR. This cooperative project was intended to maximize available resources and expertise; to serve as a model for other state archival programs; and to test the feasibility of an archival agency's reliance on an outside organization to handle materials requiring specialized technical skills and facilities.

The project had six goals:

(1) Identify MRR in selected state agencies.
(2) Prepare the Archives Division for future appraisal, accessioning, and management of these records.
(3) Train agencies' records personnel in scheduling and disposition of MRR.
(4) Develop administrative strategies to deal with confidential records.
(5) Develop a proposal for a cooperative program between the Archives Division and the University of Wisconsin-Madison.
(6) Issue reports on the project.

Information gathered through a records survey would enable the archivists to identify and describe MRR and anticipate problems of gaining custody and

transferring them from the originating agencies to the archives. This information would be used to establish appraisal guidelines and to draft disposition schedules for MRR. Training records officers and data processing personnel would be the first step in creating an effective MRR management program. Administrative guidelines and technical procedures for access and retrieval of MRR containing confidential information would assist the archives and DPLS in protecting and disseminating these records. Evaluation of the joint venture would determine whether it could serve as a model for a future cooperative program to accession, preserve, and use MR public records. The findings were designed to be shared with state agency administrators and records managers and archivists in other states.

Throughout the project, the staff were aware that their effort constituted a first step toward understanding what will be required of archivists in order to cope with new technologies. This Final Report is an effort to share experiences with other archivists who intend to develop archival programs for administering MR public records.

The Final Report summarizes the Wisconsin Survey of Machine-Readable Public Records project that was conducted between November, 1979, and January, 1981. In Part One the history of the project and the strategies employed to inventory, appraise, and accession MRR in the State of Wisconsin are described. Part Two describes the findings of the records survey. Part Three contains recommendations for establishing a MRR program for state archives. The Final Report tries to balance specificity and generality, in an effort not to burden the reader, but at the same time provide a document that will help archivists avoid the pitfalls encountered along the way. Detailed information that does not find its way into the body of the Final Report is found in the Appendices, which are available from the Historical Society.

Machine-readable records are defined as records that require access to a computer to transform their contents into a human-readable form. MRR are recorded on various physical media--punched cards and magnetic media such as tapes, disks, drums, and diskettes. Their contents range from the text of a letter, to detailed accounts of receipts and expenditures, to responses to survey questionnaires, to complex patterns of digits that represent the series of coordinates that constitute a map. In Wisconsin, MRR are public records regardless of their storage medium or contents, as long as they are made or received by a state agency in the transaction of public business (s. 16.61, Wis. Stats.).

The growth in the use of computers and the attendant increase in the volume of MRR in Wisconsin state govenment agencies make this project necessary. Machine-readable records have characteristics which pose a variety of special problems for archivists. Most machine-readable records are stored on magnetic tape, which is a fragile medium compared to paper or microform. Magnetic tape must be stored properly under stable environmental conditions and subjected to routine maintenance--cleaning, rewinding, and copying. Such maintenance prolongs the life of the storage medium and without it long-term preservation of the information on the tapes is not possible.

MRR, because of these characteristics, present new problems for archivists and records managers. They must, therefore, find ways to identify valuable MRR at an early stage in their life cycle, and then develop methods to monitor their maintenance prior to transfer to the archives. Once records in MR form are brought into archival custody, the archives must assume the responsibility and be prepared to pay the costs associated with their maintenance and preservation.

Another feature of MRR is that they can easily be updated, reformatted, copied, erased, and otherwise altered. This characteristic of the records creates additional problems for archivists and records managers. First, information can be deleted from a MR file without a trace of evidence that any changes occurred. Because of the dynamic nature of MRR, both technical and conceptual questions are raised regarding when and how to capture historical data. Unless those who design automated records keeping systems are made aware of the potential future value of the information, many systems will be designed without methods to retain information of long-term value. Second, the ease of copying and reformatting the data leads to a proliferation of records both in MR and manual form. Many of these records are closely related, yet not always identical. Archivists must identify the most complete and usable copy and determine the most desirable versions and formats in which to retain the information.

Another characteristic of automated records systems is that the MR version itself does not contain all of the information needed to access, use and understand the record. In addition to the records themselves, documentation is required that describes the contents, arrangement, codes, and technical characteristics of a MR file. Without complete and accurate documentation, archivists are unable to appraise the informational value of the files and users cannot retrieve the information or understand the file's contents.

The use of computers by state agencies has also altered the organizational environment in which records are kept. For automated systems, the responsibilities for defining, producing, maintaining, and using records are shared by three groups: the creators of records, the users, and data processing personnel. As a result, management of MRR requires coordination of the activities of all of these records keeping personnel. The archival activities of inventory, appraisal, and scheduling become more complex as well. The information needed to understand and appraise a MR file must come from a variety of sources.

Amendability is exacerbated by a prevailing (and erroneous) belief that MRR are not public records. Many records creators and custodians do not consider MR files to be "records" due to the fragility of the storage medium; the ability to erase, update, copy and reformat the data; and the short term use of the information. Unlike textual records which have a growing physical presence that demands attention, MRR are compact and the storage medium can be erased easily and reused. Because they are usually stored in a data-processing center physically removed from the offices of their legal custodians, they are easily ignored. Both their legal and physical custodians

must be made aware of the legal status of MRR and the special problems associated with them. They must then be integrated into the conventional procedures for orderly disposition of public records.

On the other hand, MRR, if properly managed and maintained, offer potential solutions to several pressing archival problems. During the last few decades, archivists have confronted an ever-growing mass of paper files. The volume of some records series is so great that it is impractical for a researcher to sort through the records manually. Some large records series must be destroyed even though they contain information which is of value for future research; archivists often find that the space required to store such records and the difficulty of manipulating and gaining access to the information they contain make retention impractical. For some voluminous records series, MRR can resolve this problem. In MR form, they are very compact, and offer the potential for rapid access, relatively easy information retrieval, and greatly increased manipulability.

MRR also have the potential of resolving some of the tension between the individual's right to privacy and the public's right to have access to information. A variety of techniques have been developed to strip personal identifiers and other identifying information from MR files. Using these methods, called disclosure-avoidance techniques, archivists can make micro-level data available for statistical research in such a way that the identities of all individuals are masked.

MRR also permit more effective research use of public records. First, the information in a MR file can be rearranged, aggregated, compared, and subjected to statistical tests without the laborious tasks of sample selection, data collection, coding, and data entry. A widely-available collection of high quality data on a variety of demographic, economic, and social characteristics of the population could significantly reduce the need for independent data collection on many subjects. Some MR files have the additional advantage of potential linkage with other files, thus providing a more comprehensive set of documentation on some subjects than would otherwise be available.

MRR can be easily duplicated and thus offer a great potential for wider distribution of research resources. In the future, researchers may no longer need to visit a central research facility because the information they need can be mailed to them or distributed through a telecommunications network. Finally, MRR can be used to generate finding aids to paper and microform files which will help archivists provide improved access to the archives' holdings.

PART ONE: HISTORY OF THE PROJECT

1. Preparation

During the first three months, Margaret L. Hedstrom, project archivist, was hired and trained. She reviewed literature about MR archives, privacy and confidentiality, research use of micro-level data, computer technology, and records surveying. She also attended a one-week, intensive training session at the Machine-Readable Records Division of the National Archives and Records Service (NARS).

The staff conducted research on state agencies' data processing activities. The State Data Processing Plan was examined; data processing personnel were interviewed; data processing and records management at the state and agency levels were evaluated; and key issues and problem areas were defined. This included defining what constituted MRR, and establishing policies and practices governing data collection, documentation, confidential information, and retention of MRR. A data collection form was developed to gather information about the contents and technical characteristics of MRR.

2. Survey and Analysis

During the next nine months, MRR surveys were conducted in the departments of Public Instruction (DPI), Revenue (DOR), and Health and Social Services (DHSS). Detailed reports of data processing activities and of the MRR series were written. (See Appendix A [DPI], Appendix B [DOR], and Appendix C [DHSS]).

In February and March 1980, a comprehensive inventory of the major data systems was carried out in the DPI. Twenty-six data systems were identified and detailed information gathered on the contents and technical characteristics of the files. Sixteen individuals were interviewed to obtain information about agency MRR policies and about specific files. Components of MRR description were defined and the format for Records Disposal Authorizations (RDAs) developed. Three RDAs for eight MR records series were drafted and approved by the Public Records Board at its April meeting.

During April and May 1980, the project archivist met with records management and data processing personnel at the DOR to explain the goals of the project, lay the groundwork for the survey, and discuss how confidential information would be handled by the agency and archives. Seventeen agency staff members were involved in meetings and interviews regarding general agency policies and specific MRR. Interviews with key agency personnel and an inventory of data files and systems revealed 13 systems which appeared to

contain data of potential research use. RDAs covering four MRR series were drafted and approved at the July meeting of the Public Records Board.

The DHSS was surveyed during May, June, and July 1980. Due to the size of the DHSS and its extensive use of computer technology for records keeping, the investigation was limited to the Division of Community Services, the Division of Economic Assistance, and the Division of Health. No attempt was made to compile a comprehensive inventory of all MR data files. Rather, the survey focused on access to confidential information, the agency's use of large online data bases, automated welfare case files, and the exchange of information between state and federal government agencies.

To carry out these three surveys, a methodology was developed to promote consistency and completeness in our information-gathering activities. The methodology described in this section represents only a general framework used to conduct surveys of data processing activities and MRR in the three agencies. Information about MRR differed among the agencies for a variety of reasons, including the degree of centralization of data processing; awareness of the research potential of the agencies' records; and the quality and availability of reports or partial inventories of the agencies' MRR. Methods became more refined as the project staff became more familiar with the problems and sources of information about MRR.

Background research into state agencies and state-wide data processing activities prepared the staff for the project and was used to develop the criteria for selecting the sample of agencies to survey. The annual Wisconsin Blue Book provided information about the administrative structure and functions of each agency, information about exchanges of data between the agency and local or federal agencies, and in some cases identification of key divisions within the agency where MRR were created. The agencies' biennial reports to the Legislature provided information about major systems development and specific automated systems. The biennial reports from the DPI made reference to several specific systems, whereas those for the DOR and the DHSS provided only general statements about data processing activities with few references to specific systems. The inventory of tapes deposited at the State Records Center and of COM center users was consulted for names of specific systems. Some RDAs contained indications that the source document or report covered was related to an automated system. The project staff interviewed key Department of Administration personnel responsible for the administration of data processing services, to gather information about implementing and operating the state data processing plan and proposed regional computing centers. Major data producing agencies were identified (i.e., agencies for which the production or use of MRR are central and integral parts of the agency's activities).

2.1 Selecting the Sample

Major data-producing agencies, defined as agencies for which the production or use of machine-readable records are central and integral parts

of the agency's activities, were identified in order to select those that would provide the project staff with an opportunity to explore a range of problems. Several sources were examined including the Wisconsin Blue Book, an inventory of computer tapes deposited at the State Records Center, an inventory of users of the state Computer Output Microforms (COM) Center, and a review of the RDAs for several agencies. Of the 56 state agencies, at least 36 produce or use machine-readable records and 17 were defined as major data-producing agencies. The University of Wisconsin, the largest single data-producing agency in the state, was excluded from the survey for practical purposes and because the University Archives is the official repository for its records. Each of the 17 major data-producing agencies were ranked according to seven criteria:

(1) Produces confidential records.
(2) Produces housekeeping records.
(3) Produces records that are likely to have research value.
(4) Has functional relationships with local, county, and federal units of government.
(5) Is dependent on data systems to meet multiple needs.
(6) Has a separate planning, evaluation and research division.
(7) Has a separate data processing unit.

The final selection of agencies was made from a ranked list and the three agencies selected met all the criteria.

2.2 Identifying Data Processing Systems and Machine-Readable Records

An effort was made to identify and compile an inventory of data processing systems and MRR. Sources for the inventory varied considerably among the three agencies. Sources included information from the biennial and internal reports listing all major computer applications, costs and products or services provided; printouts from data dictionaries; interviews with data processing, records management, and information services personnel; and already-produced guides to the agency's information system.

2.3 Identifying Information about Files

Our first step was to identify key contacts within the agency who could provide us with information about specific data systems. These key contacts included directors of data processing, information systems specialists, data coordinators and records managers. They then referred us to individuals familiar with each data file.

We found a sharp division of labor between data processors and users. It was necessary to talk with a number of individuals in order to gain a comprehensive understanding of each file. Users, who include program administrators, research analysts, and file clerks, generally are responsible for collecting the data, determining the output requirements of automated

systems, and using the system for program evaluation, planning, and reporting. Data processors include a wide range of personnel who design systems, enter data, write programs and technical documentation for the files, operate the systems, and coordinate and supervise numerous data processing applications. Some of the larger agencies have information systems specialists who serve as liaisons between users and data processors. While the latter specialists were able to provide general descriptions of automated systems and to refer project staff to the appropriate users and data processors for additional information, they often lacked detailed knowledge of the contents and technical aspects of the files.

Information about each data file was then compiled after interviews with these personnel and from published reports related to the file. One form for each data file was completed, based on the interviews and examination of source documents, RDAs for the source documents and output, and the file layout (if available). Data processors provided information about the physical structure of the data, its arrangement, retention policies, maintenance practices, updates, and master and processing files. After completing the survey in each agency, the project archivist drafted descriptions and recommended retention schedules for selected files.

2.4 Describing the Data Systems and Machine-Readable Records

The project staff initially designed two different forms for gathering information about the data systems and data files. The assumption behind the two forms was that most MRR would be in "systems" which consisted of several master files, linked together to perform a variety of functions. We found that while a description of the linkage among files is useful in a multi-file system, much of the information on the data system repeated that on the data file form. We tested the data collection form in the DPI and revised and simplified it for surveys in the DOR and the DHSS. (See Appendix D for the final form.)

The revised survey instrument was based on a form used by the NARS Machine-Readable Records Division for their 1975 survey of federal agencies' MRR. We collected three types of information about each master file:

(1) Contents and purpose of each file (elements of information that would be included in a survey of textual records). These elements included records creator, title, inclusive dates, location of the records, purpose, contents, and unit of analysis.
(2) Technical characteristics of each master file (physical and logical structure of the data, storage medium, and software and hardware used to create the file). Elements were limited to factors that would have a direct bearing on appraisal considerations (feasibility of preserving the data in its original form; costs and problems associated with transfer to the archives; potential problems researchers might encounter when using the data).

(3) Processes used to create and operate an automated system. This information allowed us to examine all the components of an automated system simultaneously and to analyze the relationships among its parts.

In an administrative environment, documentation, which is essential for interpreting the contents of MRR, is dispersed among personnel throughout the agency. To gain control over the documentation as well as the data files, a series of questions about the existence and location of the documentation was added.

2.5 Selecting Data Files

A comprehensive survey of MRR was conducted for the master file of each data system in the DPI. In the DOR, however, only data systems which appeared to contain data of potential long-term research value were selected for evaluation. In both the DOR and the DHSS, an attempt was made to select systems which would also acquaint the project staff with a variety of issues: updated files, files with multiple source documents, data collected for research, samples, large and dynamic on-line data bases, automated case files, confidentiality, and exchange of data between local, state, and federal agencies. All these issues constitute significant archival, technical, administrative, and intellectual problems.

3. Accessioning

Kathy Unertl, a member of the Data and Program Library Service staff, was hired in October, 1980, to provide technical assistance and coordinate the transfer of selected data files from the originating agencies to the archives. Negotiations between the agencies and the project staff already were underway to arrange for transfer of the data, for evaluation of the documentation, and for access to the files once they were turned over to the archives. During October, November, and December, Ms. Unertl accessioned six data sets from the DPI and the DOR. In addition, she compiled documentation and drafted finding aids for five files. Accessioning activities included locating and compiling documentation, transferring the data to new tape, verifying a printout of the data by comparing it to the tape layout, reformatting the data when needed, and creating duplicate back-up copies of the tapes.

4. Workshops

The project staff spent much of October in preparation for the workshop including developing two workbooks for the practical training sessions (see Appendix E). The objective was to disseminate the findings of the project and

provide training in records management and archival retention of MRR. The workshop for records managers, administrators, selected data processing personnel, researchers, and archivists was on November 11 and 12, 1980.

The workshop consisted of general sessions followed by practical training sessions. The general sessions, aimed at a wide range of personnel associated with MRR, were held during the morning of November 11, and were attended by approximately 150 persons. These sessions included a keynote address by Bruce Ambacher of the Machine-Readable Records Division of NARS and a panel discussion of legal issues, technology and trends, records management, archival concerns and research use of MRR. Participants in the panel included Max J. Evans, project co-director and moderator, James McDermott, assistant attorney general, Larry E. Travis, professor of computer science, UW-Madison, Mary Ann Woodke, state-wide records and forms management coordinator, Margaret L. Hedstrom, project archivist, and Martin H. David, co-principal investigator and professor of economics, UW-Madison.

Two practical training sessions in records management for MRR were held during the afternoon of November 11. The first session focused on the components of automated records systems including the computer system, the personnel, and the records. Basic data processing terminology was presented and the elements of records management for MRR were discussed. The second session provided practical training in records management for MRR. Methods to identify, describe, appraise and schedule MR data files were presented and the participants examined a case study and completed a scheduling exercise using real-life examples of MRR from state agencies. These limited-enrollment sessions were repeated on November 12. About sixty persons attended the practical training sessions representing 20 state agencies, six of the UW campuses, the City of Milwaukee, and two private businesses.

5. Reporting and Other Activities

The project staff spent December, 1980, and January, 1981, writing the final reports on the project and conducting follow-up activities. During the course of the project, members of the project staff also participated in several related activities. Ms. Hedstrom presented a paper entitled "Privacy, Computers, and Research Access to Confidential Information" at the 1980 annual meeting of the Midwest Archives Conference and a paper entitled "The Wisconsin Survey of Machine-Readable Public Records: Techniques to Inventory, Appraise and Schedule State Records" at the 43rd annual meeting of the Society of American Archivists in October, 1980. Dr. Ham gave a brief presentation about the project at the 1980 annual meeting of the National Association of State Archivists and Records Administrators in July.

Between November 1980 and February 1981, Alice Robbin analyzed federal and state statutes and administrative rules pertinent to access to confidential MRR for research and statistical purposes, examined published reports of these policies and practices, and interviewed upper- and lower-echelon agency

administrators with regard to administrative practices for these records. The results of this study are found in Technical Report #4. The Social Utility of Personal Information. An Examination of and Recommendations for Statutory Protection and State Agency Policies and Practices Regarding Research Access to Confidential Records (See Appendix F). Ms. Robbin also testified before a committee of the Wisconsin State Senate on a bill to recodify, clarify, and amplify state law concerning access to public records.

Professor David prepared a paper based on the project, "The Great Rift: Gaps Between Administrative Records and Knowledge Created through Secondary Analysis," presented at the International Conference on Computers and the Humanities, Ann Arbor, Michigan, May 28-29, 1981. Ms. Unertl conducted a study of the feasibility of transferring the MR versions of Wisconsin individual income tax returns, 1970 to date, from the DOR to the State Archives. Members of the project staff also offered guidance to records managers in several agencies regarding the scheduling of specific MRR.

PART TWO: FINDINGS

The experiences gained during the course of the project serve as a basis for generalizing about the quantity and nature of MRR in state agencies. However such generalizations must be tempered with an awareness that agencies' data processing and MRR management activities differ in the extent to which automation has been applied to records keeping activities; the quantity and nature of MRR; the sophistication of data processing methods; procedures used to manage and document MRR; and several other areas which are discussed below.

6. Quantity and Nature of Machine Readable Records

State agencies use computers for a wide range of records keeping applications. Consequently, many MRR document routine administrative activities, such as the distribution of public funds, collection of revenues and taxes, issuance of licenses, and case management of client records for state-supported or state-administered programs. Areas such as state property inventory and control, financial accounting, and licensing appear to be almost universally automated.

Many routinely-generated state and federal reports are produced with computer assistance. Automated systems are used frequently to gather and process enumerations and descriptive information about public institutions (censuses of schools, teachers, students, hospitals, health professionals, etc.). Less common applications include special studies and surveys, data collected for evaluation of policies and programs, as well as many special applications unique to each agency. The DOR and the DHSS use computers to some degree for nearly all records keeping functions. Computer applications in the DPI are less universal and are concentrated in the areas of enumerations, financial accounting, and state and federal reporting.

Agencies' use of computer technology rarely results in the collection of new types of data. Rather, computers are used to process and store the types of information that agencies have traditionally collected. Most automated systems were designed to assist in the management of large volumes of data that are subject to either frequent arithmetic manipulation or updating. While many of the files contain data that could be used for statistical analysis, information in most MR data files is ordinarily collected for administrative purposes. Thus MR public records differ in a number of ways from MR data files produced for research. The content of many files is limited to a few data elements mandated by statutes, program guidelines, and reporting requirements, or selected from more extensive textual documentation. Individuals, institutions, and businesses report information about attributes, events, and transactions on simple reporting forms rather

than on sophisticated survey instruments which solicit information on attitudes, behavior, and social or economic characteristics. Unlike survey research files, administrative files usually are part of large, multi-purpose automated records systems. They cover an entire population and are generated at regular intervals on an on-going basis.

The project staff was particularly interested in determining to what extent MR files duplicate information found in other formats. Because most MRR are extensions of manual systems, there is a strong relationship between the informational contents of MR files and the textual records. Automation of record-keeping has resulted in a proliferation of the same information in a variety of forms. The strong relationship between MRR and textual records has important implications for archival programs because records managers and archivists must analyze all the components of an automated records keeping system.

In many cases, MRR represent the core of an agency's information systems, systems which consist of both textual and MRR. Usually, the informational contents of the MR components of these systems are nearly identical to or extracted from the textual components. One result is that the same information is likely to exist in several forms (paper source documents, coding forms, magnetic tapes, COM and/or paper printouts, and published reports). Information is likely to be available as both micro-level data and as summary statistics, and in several arrangements (alphabetical, numerical, and/or geographical) with minor variations among the different forms and versions. Variations of the same information could be found in processing files, extracts from the master file, and computer printouts.

However, there are some notable and important exceptions to the strong relationships between MR files and textual records. For example, in the DOR, we examined the data base for the 1974 tax model. This file was constructed by linking several types of records containing economic and demographic data on some 20,000 households. Some of the data were obtained from MR files while other data were coded from hard-copy source documents. Thus the tax model data base constitutes the only source of such extensive documentation on this sample of households. Similarly, the DHSS maintains a large on-line data base, the Computer Reporting Network (CRN), which contains data on public assistance recipients. While more extensive documentation on these clients exists in the hard-copy case files in each county social service agency, the master file of the CRN is the only centralized source of this data for the entire state.

Because there is currently a strong relationship between the content of conventional and MRR, archivists have a short grace period in which to establish programs for long-term preservation of information in MR form. Although some such information is unique, in most cases it could be reconstructed from textual sources. However, this situation is changing rapidly as agencies begin to utilize more sophisticated computer technology and increase their use of word processors, mini-computers, on-line systems, and data base management systems (DBMS). Unlike the present situation, in

which archivists can select the most suitable records from a number of available formats, technological trends suggest that in the near future much of the documentation will exist only in MR form.

7. Problem Areas and Implications

This section of the report examines a number of problems which must be solved in developing an archival program for MRR. Since MRR must be brought under control at an early point in their life cycle, the foundation for an archival program is a solid records management program in the agencies. Thus, some of the problems discussed below must be addressed by the agencies through improved records management procedures; others by the archives through expanded programs, resources, and skills; and others through cooperation between a variety of agency, archives, research, and technical personnel. Archivists must play a leading role in stimulating interest and cooperation in this endeavor.

7.1 Identification of Systems and Master Files

The first objective in gaining control over the MRR in each agency is to gain a comprehensive overview of its records. The surveys demonstrated that agencies do not have inventories of MRR nor have they applied standard records management procedures to these records. Centralized sources of information about the agencies' MRR are either unavailable or inadequate.

Two centralized sources of information about MRR were explored, tape library listings and automated data directory listings, both of which are incomplete.

Tape library listings are computer printouts on each tape in a computer center's tape library. The computer centers produce these library listings on a regular basis to reflect the frequent changes in the tapes. While the tape library listings include all tapes in the tape library (both processing and master files), tapes located elsewhere, such as in other computer service bureaus, records centers, or users' offices, are not included, nor do they include files maintained on-line on the computer center's disk drives. The tape library listings usually provide only a limited amount of information about the tapes and use many codes and abbreviations.

The data directory listings contain more information about each file, but seldom provide enough information to identify the master files or locate the file sponsors. Wisconsin's largest regional computing center for state agencies is currently installing an automated data directory system and the DPI has a data directory for files in its data base management system. A listing from the DPI system provided some important details about the files, but could not substitute for a comprehensive inventory. It did not provide enough information about custodians and users of the files and the abbreviated

file and variable titles were too incomplete to allow for identification of the specific data sets. However, the trend among data processing centers to acquire more sophisticated software for inventory and control of MRR may eventually provide records managers and archivists with a centralized source of information about the MRR in state agencies. Elaborate data directory systems include such features as narrative and technical definitions of files, logical records and variables; automatically-produced file layouts; retention schedules; and terms of access.

The absence of a convenient, centralized source of information about agency MR data files requires that other sources be investigated. These sources vary in each agency. Where no centralized inventories are available, those conducting surveys must rely on agency reports, interviews with key data processing personnel, data processing division reports on the costs of operating major systems, guides to information systems covering special subject areas, budget requests for expansion of data processing services, and agency planning documents. Key personnel in data processing divisions and offices responsible for agency information systems can provide initial information about the titles, dates, and functions of automated systems and MR files.

Records managers, although the logical ones to provide this information, have not been involved in the inventorying of MRR. Most records managers are reluctant to initiate such procedures, usually citing lack of knowledge and experience with MRR as the reason. As long as MRR remain outside of agency records management procedures, archivists should anticipate spending considerable effort compiling inventories of MRR. This approach to identifying and locating master files is time consuming and labor intensive, yet seldom yields comprehensive results. Furthermore, the creation of new files and systems makes inventorying an on-going activity.

7.2 Revisions and Modifications of Automated Data Systems

Most MRR are created as a function of on-going administrative activities. Because most of these files span several years, they are subject to frequent modifications and revisions. Revisions and modifications are made in response to changes in program guidelines, goals, and objectives; to changes in the structure of the data when technological innovations are incorporated; or to changes in data collection, coding, data entry, and other procedures when improvements are made in the quality and efficiency of the system. Automated systems facilitate revisions of the data, new additions, error corrections, and transformation.

However, the dynamic nature of automated systems has several important implications for archival programs since the contents of one file may be different from files created by a later version of the same system:

(1) Since the contents of the file change over time, MRR may not provide comparable data that can be used for time-series analyses or other

longitudinal comparisons. In cases where the contents of the file remain relatively stable, minor changes in the definitions of variables or the scope of the file will require additional documentation to explain these changes and allow the researcher to perform the necessary manipulations to make the data comparable over time.

(2) Modifications of the technical characteristics of the file may require the agencies and/or the archives to restructure data files into comparable formats.

(3) Frequent revisions of automated systems will make inventory, appraisal, and scheduling of MRR an on-going aspect of a records management and archival retention program. When substantive modifications are made, records schedules should be revised. Because even a minor alteration to the contents of a file can have a significant impact on its research value, the records must be reappraised. For example, if the social security numbers, which previously served as the principal means of linkage with other MR files, were dropped from the file, its research value would diminish considerably.

The dynamic nature of these systems creates a variety of problems for archivists and may diminish the availability and utility of MR public records for research purposes. Although some automated systems have the capacity to store non-current data by transferring obsolete records or data elements to "history" or "archives" files, most do not maintain a historical record of transactions or retain data on closed cases. When the status or characteristics of a case change, current information replaces obsolete data and the MR form of the historical record is lost. Other systems have no method to identify and purge closed cases or to distinguish closed cases from active ones. Still others are designed to retain only selected data elements from the active files when cases are closed.

For archivists, the basic problems with automated systems are the loss of historical information and the difficulty of determining when to accession data from such a dynamic environment. One strategy for capturing data from systems that do not generate history files is to create periodic "snapshots" of the master file, that is, to make copies of the MR file at specified points in time. While such a strategy would not provide a complete historical record, it might provide an acceptable statistical profile of the population at regular intervals. The feasibility of this approach depends on the size of the data base, frequency and extent of updates, the structure of the data, and subject matter covered by the records.

7.3 Maintenance and Preservation

Machine-readable records are stored on magnetic tape, disks, drums, or punched cards, all of which are fragile in comparison to traditional paper and microform records. Unlike paper and microforms which can be placed in inactive storage for relatively long time periods (50 to 100 years) without

loss of information due to deterioration of the medium, MRR require regular maintenance to assure preservation. In addition, if data in MR form are stored on COM or paper, the information would have to be reconverted into a MR format in order to carry out statistical analysis on the computer. At present, most state agencies maintain their data files on magnetic tape, which is assumed to be the most efficient and economical way of storing inactive files. However, magnetic tape is a fragile medium, requiring careful and regular maintenance to ensure its preservation.

MRR stored on magnetic tape, disks, and drums pose different archival problems. Although from one point of view, they promote great efficiencies because of their capacities for large-scale storage and rapid access, they cannot be easily transferred from the originating agency to the archives. (This point is discussed in more detail in the section on Hardware and Software, pp. 28-30.)

In state agencies, responsibility for MRR maintenance and preservation is left to the user. The data processing center provides these services at the user's request. Consequently, the amount of attention paid to maintenance and preservation varies considerably among files.

Although many users and some data processors are unaware of basic maintenance requirements and procedures, maintenance problems with active records appear to be minimal because the files are used at least once a year for updates and revisions. However, inactive files often are neglected. In cases where agencies have assigned relatively long retention periods (5 years or more) to MRR series, little or no consideration is given to basic maintenance requirements which will assure that the records will be readable throughout the retention period.

The preliminary inventory of MRR deposited at the State Records Center revealed numerous files deposited there since the early 1970s, that had not been recalled by their custodians for maintenance or use. Agencies have not developed records keeping procedures for the tapes on deposit at the Records Center and most files there are not scheduled. Some records have future potential applications, but the absence of maintenance may render them unusable. Furthermore, documentation is not transferred to the Records Center with the MR files, making future access to the records impossible in most cases.

Besides physical problems which potentially might develop with magnetic tapes, fundamental changes in computer technology may result in the obsolescence of the storage format. Specifically, the hardware and/or software used to generate a tape may be phased out entirely. Software-dependent files may be unusable if the programs and computer operating systems necessary to access the data are no longer maintained, unless provisions are made to update files as the software and hardware change. These provisions are unlikely to be made for the vast majority of records in inactive storage at records centers.

Maintenance of MRR has a low priority in the agencies. Many data processors and users are unaware of the rationale for and methods of tape maintenance. Consequently, archivists must take an active role in the education and training of agency personnel in this area. They must also be prepared to accession (or copy) MRR after a much shorter retention period than for manual files.

7.4 Documentation[1]

Three recurring problems associated with documentation were identified during the course of the survey and merit discussion: organization, quality, and maintenance. Factors which influence the organization, quality and maintenance include:

(1) The agency's ability to establish and enforce formal documentation.
(2) The extent to which a file has multiple purposes and multiple users.
(3) Anticipated and actual research applications for the data.
(4) The idiosyncracies and needs of programmers, administrators, and users of the data.

The organization of documentation reflects the division of labor in automated systems. Data file documentation is fragmented and dispersed among administrative and data processing personnel in the agencies. Portions of the data file documentation pertaining to technical characteristics of the file, are usually kept by the data processing staff. Creators and users of the records also keep important elements of the descriptive documentation.

Archivists must focus on the documentation which describes the contents, arrangement, and technical characteristics of a MR data file. Of secondary importance is the systems documentation which can provide valuable insights into the relationships between different components of an automated system and may be applicable to identification and evaluation of related MR and textual records. Program documentation, on the other hand, is applicable almost exclusively to the daily operation of the system. Often, however, portions of the data file documentation is interfiled with systems and program documentation. Thus, in order to evaluate the documentation for a MR file, the archivist first must identify and select the relevant portions from a much larger set.

Most agencies agree that the quality of documentation must be improved to satisfy the agencies' own needs for information. Agency efforts to improve

---

[1]Documentation refers to the descriptive information about the operation of a system and relationships among the hardware, software and data (systems documentation); software instructions (program documentation); and arrangement, content, and coding of the data (data documentation or codebook).

the quality and scope of documentation are concentrated, however, on current and future systems rather than on systems that are no longer active. Documentation, if available at all, often suffers from several problems. The documentation is sometimes unusable by anyone who is unfamiliar with the system because abbreviations, unexplained codes, or illegible handwriting are used. Furthermore, some of the documentation may be inadequate because users and file sponsors, who are familiar with the meaning of each variable, do not bother to formally explain what, to them, is obvious. Frequently, both the technical and descriptive documentation are lost because they are not written down. If file sponsors and technical support staff are no longer at the agency, no documentation may exist.

A related problem is maintenance of documentation and results from the frequent revisions of the automated records system; modifications and changes to the data frequently are not recorded in the documentation. When a data file is retired from regular use, the documentation for previous versions of the file is seldom compiled and maintained. Thus there are numerous MR files (especially from systems that have been revised and those prior to the mid-1970s) for which documentation cannot be located. This problem is exacerbated by a high rate of turnover among data processing personnel and records creators which hinders efforts to reconstruct documentation for older files.

The results of the survey indicate that the absence, inadequacy, or inaccuracy of documentation may make some data files unusable despite the apparent value of the records. Implementation of records management programs in the agencies, which must include careful attention to the issue of documentation, could help solve this problem. It could also reduce the amount of effort required of archivists while improving agency information systems by facilitating use and transportability of data in the agencies' custody.

7.5 Retention

Retention periods for MRR, if set at all, are set on a case-by-case basis and usually are determined by the records creators and data processing personnel based on perceived administrative and legal requirements for retention of the data. These practices usually exist outside of the prescribed procedures for proper disposition of public records.[2]

---

[2]The Wisconsin state public records laws require that each new record series established by a state agency be "scheduled" within one year of its creation. This schedule, which briefly describes the records and includes a recommendation for ultimate disposition, must be approved by the Public Records Board, which consists of a representative of the Governor, the Attorney General, the State Auditor, and the Historical Society.

Although retention periods have been fixed for some MR files, they are almost always determined through internal agency procedures without Public Records Board review. The issue of retention tends to receive more attention in cases where MR files represent the only source for a particular body of information. (In the case of special studies and surveys, the retention period for the MR files often coincides with completion of the project.) In most cases, however, no fixed retention periods have been established. A common agency practice is to periodically evaluate the agency's tape library. During such an evaluation, inactive tapes are identified and the records creators are asked to specify which files can be "scratched." If the records creator grants permission, the tapes are recycled.

Many files are transferred to the State Records Center for off-site storage either as inactive records, or more likely, as security back-up copies of master files. Because such files are not scheduled, they are often forgotten. In November 1979, approximately 15,000 reels of computer tape, including both inactive files and security back-up copies of current records, were on deposit there.

7.6 Scheduling

Procedures for the scheduling of MRR differ from those for textual records in several ways. The vast majority of MRR are temporary "processing" or "work" files, used to create, revise, rearrange, and back-up the more permanent master files (containing data in its most consistent, organized, and accurate form). Scheduling efforts must concentrate on potentially valuable master files which often constitute the core of an automated information system.

Given the potential research value of many administrative data sets, evaluation of each master file is preferable to application of a general schedule which allows for destruction of certain types of master files. This approach deviates from the procedures used by NARS through use of its General Records Schedule (GRS) 20 for MR files. Several of the GRS 20 categories of master files did not correspond to the types of MRR created by state agencies. However, it is possible that further research could lead to the development of general schedules for some types of state records in MR form, such as licensing files and fiscal master files or to general schedules for all records associated with particular types of systems. Public records legislation does not differentiate between processing and master files. Therefore, methods must be developed to allow data processors to "scratch" and recycle processing files when they cease to have current uses. This could be accomplished through the development of a general schedule for processing files, which would grant records creators, data processors, and records managers some discretion over the retention of processing files. The general schedule could relieve the agencies of responsibility for writing retention schedules for these ephemeral materials and allow records managers and archivists to concentrate on the systematic scheduling and appraisal of potentially valuable master files.

The relationship between textual and MRR in state agencies suggests that scheduling should be integrated with existing programs for textual records. Comprehensive schedules can be written to cover all the records associated with an automated system regardless of physical medium. These records would include paper source documents, coding forms, reports, and other paper and COM printouts; and MR processing, master and extract files. The retention periods will vary for each component of the system, but the comprehensive schedule will assist records managers, records custodians, and users of the records to retain the most useful versions. This comprehensive approach to scheduling will also help the archivist to gain an overview of an integrated records system and to select the best formats and versions for long-term retention.

Ideally, MR and related textual records series should be scheduled while the system is in the design stage or as soon as possible after the records are created. Through prompt scheduling, MRR that merit long-term retention can be identified at an early stage in their life cycle and measures can be taken to assure proper maintenance of both the physical medium and the documentation. Such schedules can motivate agencies to evaluate the administrative, legal, and research needs for access to large bodies of information and to select the most appropriate versions for retention. Early scheduling will also assist data processing personnel in integrating retention schedules and data transfers into their normal operations.

Current practices for MRR scheduling are not adequate to meet the goals of improving the efficiency of agency operations and of identifying historically valuable records for archival retention. Efforts to reduce the inventory of inactive data files in tape libraries result in hasty and arbitrary retention decisions. Decisions to "scratch" tapes because the file sponsor is unaware of any future applications outside of his program area can result in destruction of records for which potential agency and external research interests exist. The lack of procedures for orderly destruction of inactive records results in significant quantities of computer tape and other storage media being used for data that lack any future applications. And, in some cases, by the time a decision is made to "scratch" a tape, the physical medium has deteriorated to such an extent that it cannot be reused.

An archival program for MRR retention is dependent upon solid records management practices including timely and comprehensive scheduling of all types of records associated with automated systems. Yet to implement such records management programs, both records management and data processing personnel need education and training about the importance of and techniques for scheduling MRR. While archivists can provide valuable guidance to agency personnel in these areas, ultimate responsibility for scheduling lies with the personnel in the agencies who are familiar with these systems.

7.7 Appraisal

The methodology for appraising MRR is the same as for manual records and goes hand-in-hand with scheduling. The schedules specify the period of time

the records should be retained in the originating agency and include a recommendation regarding the ultimate disposition. This recommendation is reviewed by archivists to determine whether the records have sufficient value to justify their transfer to archival custody. Appraisal techniques should be based on basic principles and procedures developed and practiced by the Machine-Readable Records Division of NARS, social science data archives, and appraisal principles applied to manual records.

Because most MRR are created for administrative and not analysis purposes, they are like manual records in that they must be appraised for their secondary research value. MRR may be created for very specialized purposes and may concern a small segment of the population or document a specific type of activity. In these cases the appraisal archivist may find it necessary to consult researchers who are familiar with the methods, sources and research trends in specialized fields.

An important aspect of appraisal is determining the appropriate form or forms in which to retain a specific body of information. Often alternative versions of MRR exist in hard-copy; it becomes necessary to appraise both versions to determine the most appropriate form to retain. For example, files that are likely to be used for reference to a single case would be most usable in a manual form. Files that are likely to be used for statistical analysis or for describing an entire population or a subgroup are most valuable in MR form. Some files are likely to be used both ways, making it necessary to retain both manual and MR versions. Archivists must also take administrative and reference uses of MRR into account. Some data files lack sufficient detail to merit permanent retention for statistical analyses, but can be used to create indexes to paper and microform files, to develop sampling frames for analyses of related files, for sample selection, or for linkage of related records.

More often, the MR and textual components of records series will not have identical contents. For example, if the output of a system consists of reports containing summary statistics, the source documents and MR master file will be the only available versions of the micro-level data. In these cases the archivist must decide whether the summary statistics generated by the agencies are adequate for description and analysis or if the micro-level data can be used for additional analyses.

When the source documents represent an alternative to the information in a MR file, the archivist must weigh the costs to the archives of preserving the MRR against the cost to researchers of recoding and re-entering data from the original source documents. Another cost that must be considered in the appraisal process is acquiring data from the originating agencies and transforming it into a format that can be used by researchers.

More extensive descriptive information is required for MRR appraisal than for textual records. In some cases, the presence or absence of a single data element may influence a decision to preserve or destroy a MR file. Essential are a complete list of the file's contents; an understanding of how the file

is generated; and an understanding of problems of accuracy, reliability, and validity of the data. In addition, appraisal decisions must consider technical factors such as the storage medium, the structure of the data, and the size and complexity of the file, all of which will have important implications for the costs of acquiring and preserving the data.

The potential research value may not always be apparent from descriptions of the file on the RDA. Furthermore, the survey revealed that the descriptive information necessary for appraisal ordinarily is not readily available. Additional training of records managers in the basic descriptive elements for MRR, encouragement of systematic compilation of documentation, and systematic inventory and scheduling activities by records management and data processing personnel will result in descriptive information upon which appraisal can be made.

It is especially important that appraisal of MRR occur as early as possible in the life cycle of the records. If the files that eventually will be transferred to the archives are identified shortly after their creation, special attention can be paid to their preservation and maintenance as well as to locating and improving the documentation. In addition, if problems arise in archival accessioning, processing, or using the records, archivists and researchers can return to the records creators for additional information and technical assistance. Files that do not merit long-term retention can be scheduled for disposition as soon as they cease to have administrative value to the agency.

7.8 Hardware and Software

Operating systems and data base management systems (DBMS) examined during the project have evolved since they were first installed or created because more efficient ways of using the technology have been discovered. In some cases, software has become obsolete and has been replaced by a new software system. Data bases embedded in one DBMS may be redesigned for new software. Data are software dependent, embedded in programs designed to structure and retrieve data elements according to predesignated applications and products. Software often requires a specific hardware configuration, thus making the data hardware dependent as well. Without a compatible operating system and applications software or the production of independent data files, again requiring software and advance planning, these data bases cannot be moved from the originating computer. Alternatively, they may be moveable, but only after a significant programming effort.

Agency use determines the design of and applications for the DBMS. Systems are designed to be consistent with an agency's information delivery, administrative, and regulatory responsibilities. As a result, there are few incentives to consider other or future information needs, particularly those unassociated with the agency's mission. The archivist's mission of creating a historical record of agency activities for future investigation will rarely coincide with the agency's programmatic agenda. Intellectual and technical

problems usually occur when the contents of these information management systems are transferred to other environments and different software are applied to the original data bases.

We have already noted (section on Revisions and Modifications of Automated Data Systems, pp. 21-22) that there may be no facility for maintaining a historical record. The data base management system may be designed to capture only current information. Thus future empirical investigations may be hampered by the inability to carry out certain kinds of longitudinal research because historical data are missing or because linkages between historical and current information are not provided.

Statistical software used by social researchers is not designed to access or retrieve and manipulate the data in the same way as information management systems. Data are imbedded in systems and applications software is designed to improve an agency's efficiency in locating a case record and linking that record to pertinent agency data. Performing statistical manipulations of small subsets of information (either data elements or cases) as is typically done by researchers is usually a secondary activity, handled by "report writing" software with limited capabilities. Further, within the data base, different files are linked in a complex set of relationships. These relationships are defined at the time the data base is constructed. Removing the different files from their "DBMS environment" disturbs these relationships (although they can be recreated at a later date). Yet, because of the dynamic development of the DBMS and its supporting computer hardware and software, these files must be removed from the DBMS to assure retrieval for statistical purposes.

Retaining all the data from these large data systems is not possible. The technology supporting agency applications imposes constraints on the structure of and access to the records. Thus, the archives must decide what portion of the record of the agency's activities is worthy of retention for future historical research and when to intervene in producing an archival record.

Complexity of the DBMS structure and its contents requires a far more sophisticated set of methodological, substantive, and technical tools than most archivists have at present. Creation of an archival record entails agency cooperation on a scale hitherto unnecessary. Assessing the utility of the data base for future scholarly applications will also require a significant degree of experience in social research, and will depend on access to social researchers, statisticians, and computer specialists. Extracting an archival copy of the data will also require a new set of costs which have not been associated with archival activity.

The archival implications of the increasing use of specialized software by state agencies are profound. Increasingly larger segments of the historical record will be lost unless archival agencies develop the technical capacity and skills needed to restructure data files into a format that can be used outside of the originating agencies. Most software dependent data files can be converted to a format that does not require use of a particular set of

software, but such conversions can be costly. The archives and/or the originating agencies can expect to incur expenses for both computer and analysts' time. Furthermore, as archivists become involved in restructuring data and capturing records from DBMS, they must be prepared to reexamine the relevance of the principle of provenance and to exert more influence over both the contents and structure of the records that will be retained.

Archivists must be prepared to deal with the prospects of accelerated changes in hardware, software, and storage media as new technological innovations become available. These innovations could resolve some current problems with MRR. For example, the development of more compact, stable, economical, and interchangeable storage media would resolve many of the current preservation problems. Similarly, efforts underway to design methods for interchange of data among software systems could reduce the need for reformatting software dependent files. While technological change will eliminate certain problems, it will also create new ones for the archivist; it will require that the archivist be aware of technological changes and make provisions for data transfer to the most current generation of storage media, before technological advancements make the archival holdings unreadable.

7.9 Accessioning and Processing

One of the objectives of the project was to accession and process several MR data files. It was necessary to determine what is involved in the transfer of these records to archival custody, and what is required in terms of cost, staff time, and expertise to prepare MRR for research use. The procedures followed in accessioning were similar to those used by the U. S. National Archives, the Public Archives of Canada, the British Public Records Office, the Data and Program Library Service, and other social science data archives.

The project staff was required to transform data files to a standard physical format on magnetic tape and to compile documentation. To do this, it was necessary to identify some key technical information, such as the physical characteristics of the data and the recording specifications of the magnetic tape. A standardized magnetic tape information form was used to gather this technical information from the agencies. The ability of agency personnel to supply this basic technical information varied considerably. Sometimes complete technical specifications and information about problems with the storage medium were provided. In other cases, only partial or no information was available. In most cases information on the recording characteristics was made available.

While it was possible, with some detective work, to determine the basic technical information if the agency failed to provide it, far more time and effort were required on the part of the archives' data processing specialist than might have been necessary under ideal conditions. For example, this was especially noticeable with the magnetic tape which arrived among the textual records of the CETA and Wisconsin Women Project, a research project conducted by the Governor's Commission on the Status of Women. (These records came to

the Archives independently of the survey of MR public records.) A sheet of paper accompanying the tape contained only a brief identification of the data file contents and the UNIVAC-dependent utility software used to write each data file. All other information had to be determined.

A variety of techniques were used to identify all the technical information necessary to complete the accessioning. For example, if the recording mode of the magnetic tape was not specified, any accompanying printouts of the records were examined for additional information. If this procedure, along with knowledge of the normal data processing activities of the creating agency, did not clearly identify the mode, specialized utility programs were used to ascertain the characteristics of the file. Additionally, because most of the tapes contained IBM standard labels, it was possible to determine the number of blocks from information in the header label. If this was impossible, blocks of each data file were printed and a count of the number of physical blocks was generated. Once the total number of blocks, logical record size, and blocking factor were known, the number of records could be determined. In addition, available published statistics were checked to verify the number of logical records.

The MRR processed during the project arrived in a variety of software and hardware dependent formats. Four of the five files were in IBM packed decimal format. Seven files of the CETA and Wisconsin Women Project included both UNIVAC software dependent raw data files and SPSS systems files; another file was in BCD character code.

Once the technical specifications were determined and available accompanying descriptive documentation accessioned, a variety of processing procedures were followed to produce an archival master copy and security back-up copy of the data in a standardized recording format (EBCDIC or ASCII). All of the data files were written on new magnetic tapes and the original agency tapes were either returned or discarded. Files were written in a standard transfer format. We printed out selected records, which were then compared with the record layout, codebook specifications, and published statistics. Typically, there were some discrepancies between the record layouts and the data, so source documents were relied upon to resolve the descrepancies. Inconsistencies between the data and documentation or source documents were noted in the user's guide produced for each data file.

One of the most time-consuming aspects of the accessioning process was deciphering the record layout which serves as the key to the location of each item of information. Some of the typical problems with record layouts included discrepancies between source documents, data, and record layout; illegible formats and unclear specifications of decimal variables; and the use of brief abbreviations for variable names.

While the record layout meets immediate agency needs to document the data, anyone not directly involved in creating the data file might have difficulty deciphering the information. Furthermore, unless an effort is made to preserve the old record layouts when major changes occur in a system or when a

system is no longer in operation, these data files may not be accessible.[3] This is especially true for data files with multiple sources of input, with derived data elements, or with unavailable source documents. It is sometimes possible to recreate documentation and record layouts if the same documents are available.[4]

A user's guide was created for each MR data file. The user's guide is similar to traditional archival finding aids, such as registers or inventories; it serves as the source of information for interpreting and accessing the data. (See Appendix G for an example of a user's guide.) The user's guide includes an abstract which provides the user with core information to determine whether further examination of the data file is warranted. In addition, including a printout of several records assists the user in understanding the structure of the data. A copy of the source document helps determine whether all the items have been converted to MR form. The source document typically contains information on definitions of terms for the data elements, and on relevant statutes. Bibliographic control was applied, with standard title pages and catalog entries for the user's guide and for the machine-readable data file.

Another key component of a user's guide is a codebook, which defines the values represented by the data. Codes for files which contained primarily administrative statistical data with only a few standardized identification fields were typically assembled from agency coding manuals. In some cases, where published code manuals were not available, codebooks were created by the archivist from other available sources. Often the source documents include a list of codes and the values for selected variables.

The various costs of accessioning the MRR include the purchase of new magnetic tapes, computer time, and personnel time. While the cost of a magnetic tape is fixed, the computer costs and analyst's time vary considerably. The computer costs ranged from $3.00 to $30.00 to produce a master and back-up copy of a data file on tape, but these costs are trivial in view of the extensive amount of staff time required to understand the files and produce a usable archival copy.

---

[3]For example, since apparently none of the original government or university researchers preserved copies of the documentation for the Individual Income Tax Return Sample, 1963-1966, this valuable research file probably will be unusable.

[4]It was possible to reconstruct the file layout for the DPI's Ethnic Data File for the 1974-75 school year. Since the information in the MR record closely followed the information contained on the source document, it was possible to salvage this data file by examining dumps of the data, more recent record layouts, a blank source document, and actual information from selected schools.

7.10 Access

Archival accessioning and processing of MR public records is aimed at making these records accessible to research. While computer technology can facilitate access to and retrieval of information, it also creates practical and technical barriers to access by outsiders. Problems of access occur in two situations: one, when the records are still in agency custody, and two, after transfer to the archives.

Use by the public of MRR in agency custody is very difficult. One of the major problems identified by this project is the absence of readily available information about MRR contents and their location in the agencies. Efforts to compile comprehensive inventories of MRR in state agencies, even after acquiring considerable experience in identifying and locating MRR, did not uncover them all.

Closely related to the problem of identification is the need for technical expertise and for access to computing facilities in order to use the records. Because these records cannot be identified visually, the requesting party is dependent upon agency personnel to specify whether or not the file exists, which technical measures are needed to gain access, and what costs might be encountered. Some agencies have defined procedures for access to their MRR by outside parties, including whether or not copies of the data will be provided, how requests for special statistical computations will be handled, and who must pay for the costs incurred. Other agencies have not addressed these issues. Consequently, the lack of general policies for access can lead to contradictory conditions for access, blanket denials of access, and access only if certain conditions are met.

Some agency personnel are reluctant to provide direct access to MR files and prefer to provide hard copies or to perform statistical analysis at the user's request. The reluctance to provide direct access can be explained in part by reservations on the part of records custodians about the accuracy and reliability of the data, coupled with their concern about the possibility that the records will be copied and redistributed without authorization.

Confidentiality is another barrier to access. Many MRR contain confidential information to which access is restricted by statutes and administrative rules. In these cases, denial of access is based on formal regulations and not on consideration of the format of the record. In other cases, there is reluctance to provide access to non-restricted materials simply because they are in MR form. This reluctance can be reduced by stressing that MRR are covered by legislation that governs access to public records.

Agency personnel are also reluctant to transfer to the archives records containing confidential information with their personal identifiers. But because personnel identifiers have potential research use as a basis for linking records on the same individuals from several sources, the archives should make every effort to obtain a complete version of such files even

though it cannot release the records in a form that would allow individual identification. This suggestion is based on a similiar policy for manual records covered by statutory restrictions; they are often accessioned, even though they can be used only under restricted conditions.

Problems of access to records in the archives are not attitudinal problems, because archivists are committed to making their records as freely available as possible. They are limited, however, by both technical and legal constraints. But one of the advantages of MRR is that they offer the potential for resolving some of the tension between the individual's right to privacy and the public's right to have access to information. A variety of techniques have been developed to delete or temporarily suppress personal identifiers and other identifying information from MR files without altering the remaining data in the files. Using these techniques, the archives can make portions of confidential records available for research in such a way that the identities of individuals are masked. To take advantage of this valuable feature of MRR, archives must develop skills to accession and maintain MRR and to perform the necessary technical and statistical procedures to create public use versions of restricted files.

Implementation of records management and archival programs can improve access to these MRR. An inventory and scheduling program will provide basic descriptive information about current records. Archival review of disposition schedules will reduce arbitrary decisions to destroy files for which research applications may exist. Dissemination of descriptive information about the records, compilation of accurate documentation, and technical assistance for researchers are long-range goals of an archival program which will greatly improve access to these materials. The creation of public use versions of files containing confidential information make research possible while still protecting individual identities.

7.11 Confidentiality

During the project Alice Robbin examined the extent to which federal and state statutory protection exists for research access to individually identifiable records produced by the three state agencies. It was determined that limited protection exists for scholarly access; policies and systematic procedures are lacking, and record-keeping practices impede both agency and researcher access.

Legislation contributes to whether scholarly access to government records is facilitated or impeded. Although Wisconsin has an excellent open records law and there is much good will among agency administrators and a desire to accommodate research needs for data, Wisconsin's statutes offer little protection for scholarly research. (See Appendix F for recommendations for modifying agency access and use policies and practices.)

7.12 State-Federal Data Transfer

During the last 15 years, numerous governmental programs have emerged that require close cooperation between federal and state governments. Many revenue sharing programs depend on local and state statistical information for determining eligibility and funds. The project staff examined the cooperation between the state and federal government for collection, processing, and transfer of state-generated MRR.

Some of the statistical information that later gets transferred to various agencies in the federal government is produced as MRR and is transferred in either MR or printed form. Increasingly, as standards for data transfer are formalized, more records originating as MRR will be transferred in this form to the federal government.

The only agency for which this question was closely examined was the DHSS. Based on that examination, there is some evidence of differences between access conditions established by the state and the federal governments. In some cases, the state agency imposed more stringent safeguards on data than the federal government, and the state agency-produced MRR had more detailed information than the same file transferred to the federal agency. It was also found that different retention policies governed the disposition of the data. For NARS to retain these state-produced records would require negotiating an inter-agency agreement because according to some contracts, the MRR are the responsibility of the state.

It is clear from conversations with agency administrators at the federal and state levels that there will be increased state-federal government data transfer in the future. Archivists will have to investigate these data transfers and contracts and identify custodial responsibility and access conditions in order to make appropriate appraisal and retention decisions.

8. General Issues

In addition to the specific problem areas which relate to a MRR program, there are three issues of a general nature that need some elaboration: agency attitudes about MRR, staff expertise, and costs.

8.1 Attitudes about the Importance and Feasibility of an Archival Program for Machine-Readable Records

Agency administrative, records management, and data processing personnel are not yet convinced that MRR are public records and thus subject to the same access, maintenance, and disposition requirements that apply to all other public records. Agency personnel rarely perceive MRR as records because they are used to create, update, and revise more permanent non-MR files. Because different versions of much of the available MR information exists in

hard-copy, agency personnel often regard the MR version as a duplicate or non-record copy.

Agency personnel must be convinced of the importance of their MRR for current and future research. They see the records as pertaining to routine administrative activities and as lacking information content that merits long-term retention. In addition, data processors and MRR users frequently have reservations about the accuracy, reliability, and quality of their data. Agencies generally do not consider MRR preservation a high priority. As more records are available only in MR form, however, agency personnel may become convinced of the importance of their records.

Many agency records management and data processing personnel do not recognize the value of a MRR management program. Most agency personnel are unaware of cases where tape files were "scratched" inadvertently and where documentation is inadequate for retrospective analysis. MRR are stored in tape libraries removed from the offices of the legal custodians of the data. MRR lack the growing physical presence of paper records which makes the need for records management obvious. Many MR files can be recycled and the physical medium reused. Furthermore, most agencies have not yet realized that clear identification of MRR through inventory and scheduling can facilitate data sharing between agencies and often reduce redundant data collection.

Agency personnel are skeptical about the feasibility of long-term retention and non-administrative use of MRR. Many administrative and data processing personnel argue that the difficulties of data transfer and long-term preservation are insurmountable in an era of rapid technological change. Others argue that without participation in design, data collection, processing, and use of these records, researchers are unable to understand the file contents well enough to interpret the data accurately. While there are reasons for concern over some of these issues, archivists can counter these arguments by pointing to examples of archival preservation of MRR[5] and by exhibiting an awareness of the interpretive problems associated with MRR. In the meantime archivists must apply strict appraisal criteria to MRR and their documentation, work with agency personnel to identify unique, high-quality, permanently valuable files, and improve the documentation for them.

An attitudinal problem common among both non-technical agency personnel and archivists is a reluctance to deal with MRR. The lack of skills, knowledge, and techniques for handling these records are often cited as reasons for this reluctance. In addition, agency personnel have difficulty determining where to begin to gain control over these records. Initial identification both of the files and of the key personnel presents obstacles, as does a lack of familiarity with computer terminology. Stimulation of

---

[5]In particular NARS, the Public Archives of Canada, and social science data archives.

interest and concern among records managers and archivists will require cautious guidance and encouragement and demonstration of the feasibility and benefits of records management and archival programs for these records.

8.2 Staff Expertise

The processes used to create MRR require the skills of a wide range of specialists who become involved in what once was a unified process of record creation and maintenance. Records management personnel should be involved in the life cycle of MRR, but frequently are not. Several records management functions are not usually being implemented for MRR. These include inventorying and scheduling, monitoring maintenance, and compiling documentation in a centralized location.

A MRR archival program must rely on agency records managers to play a crucial coordinative role. This role includes identifying the creators, users, and data processing personnel responsible for major systems; evaluating administrative and research applications for the data both within and outside the user division; and overseeing the physical medium and documentation. To carry out this role, records managers need training in identifying, describing, and scheduling MRR; in the fundamentals of automated systems; and in communicating with data processors and administrators about MRR management problems.

Training agency records managers will reduce the participation required of the archives staff in conducting inventory and survey activities and in scheduling records. The archives staff will then be able to concentrate on appraisal, accessioning, processing, and maintenance of MRR. But to do so, archivists must also understand computer systems, be familiar with the principles of documentation and appraisal, and understand the technical and descriptive requirements for transfer and use of MRR.

Most archivists have not had extensive exposure to computers. Yet elementary computer skills are needed for the basic accessioning activities of copying tapes, verification of the data, and tape maintenance. More sophisticated technical skills are needed for reformatting files, extraction of data from data bases, creating disclosure-free public use versions, and transforming data into a software independent format.

There are three alternatives for obtaining the technical skills needed for a MRR program. One is to train existing archives staff. Another is to recruit archivists with some knowledge of automated records systems and computer programming. A third alternative is to subcontract with free-lance programmers, computer service bureaus, or agency data processing staff for the required technical services. Each archives will have to select one or a combination of these alternatives, based on its personnel and financial resources and the complexity of the data being accessioned and processed.

There are particular advantages to utilizing agency data processing personnel for resolving technical problems associated with particular files and systems. The agency data processing personnel are familiar with their own systems, especially those with unique custom-built software. The disadvantage is that archival-related activities are a low priority in the agencies and the archives might experience considerable delays in transferring of data to the archives. Furthermore, the issue of who should pay for these data processing services remains unresolved.

8.3 Costs

Archivists must realize that there will be new costs associated with acquiring, processing, maintaining, and providing access to MRR.

Based on an assessment of tape maintenance procedures, it appears that MRR should be transferred to new tape when they are transferred to the archives. Thus the cost of the tape will be one that the archives must bear in order to assure that the data are maintained on a high quality medium. Expenditures for computer tape would replace costs for archival supplies such as acid-free boxes and folders, microfilming, and other preservation materials.

A second cost is analysts' time. If the archives staff includes a competent technical expert, some of the costs for analysts would be absorbed. In most cases, however, the archives will have to seek outside assistance from analysts in the agencies who are familiar with the design and operation of specific systems or from computer consultants. The costs of the analysts' time will depend on the complexity of the files and the types of transformations necessary to make the data transportable. Additional costs will be incurred if the archives creates public use versions of files containing confidential information or uses MRR to create finding aids for hard-copy files.

Another cost component is the computer time needed for copying and reformatting data files. This expense is minimal except for large, complex transformations of the data. Some costs will also be incurred for the maintenance and preservation of the MRR, including the expense of renting or maintaining environmentally controlled storage areas, as well as minor expenditures for tape maintenance procedures. Copying and preparing the documentation are also expenditures that must be borne by the archival agency.

However, archivists should also realize that there are tradeoffs. The costs for analysts' time to reformat files replace the processing activities of arrangement of a textual records series. Much of the information needed for adequate description of MR files can be obtained during the appraisal process, and additional descriptive information can be gleaned from the documentation. Furthermore, the costs of reproduction of archival materials in MR form are minimal compared to those of photocopying or microfilming paper records.

One remaining issue is how costs for a MRR program should be distributed among the originating agencies, the archives and researchers. The project staff did not resolve this issue. However, areas where activities and costs for MRR are parallel to those for conventional records were identified. For example, the archives assumes the cost of arrangement and description of paper records, a process which can be considered parallel to copying tapes as they are accessioned, reformatting data files to make them available for use,and compiling the documentation. Similarly, as researchers are expected to bear the costs of photoreproduction of paper records, they would be expected to cover the costs of copying tapes for their research purposes. The problem arises when MRR require new expenditures, particularly for outside analysts' time to extract data from a database or restructure complex files before the data can be accessioned by the archives. Currently, such situations are negotiated on a case-by-case basis. Policies and procedures for distributing these costs must be developed once more experience is gained in this area.

PART THREE: AN ARCHIVAL PROGRAM FOR MACHINE-READABLE RECORDS

9. Requisites for a Machine-Readable-Records Program

This part of the report describes the requisites and presents the elements of a program for MR public records. The requisites include a records management program which incorporates MRR; archival capabilities to handle MRR; access to outside technical resources; and legislative and administrative guidelines to govern access to confidential MRR.

9.1 A Records Management Program

MRR are unstable, updatable, and stored on a fragile medium. They must be identified and controlled at an early stage in their life cycle. Thus the quality of MRR archival programs is dependent upon the procedures used in state agencies for handling these materials. A MR public records program grows from the records management procedures in the agencies. Such programs can be patterned after existing inventory and scheduling procedures for textual records, or modified where necessary to account for special technical considerations. MRR must be incorporated into formal disposition procedures which permit no destruction or transfer of public records without archival review and approval.

Records management program objectives must be:

(1) Educate records managers and data processors about the importance and legal status of MRR.
(2) Provide training in the techniques needed to inventory and schedule the records.
(3) Develop guidelines for writing and maintaining documentation for files with archival value.
(4) Assure that data files in agency custody are maintained and preserved.

State archival agencies must take the initiative in encouraging the establishment of this program. There must be, in addition, support from central records management officers and a commitment to incorporate MRR into existing records management procedures.

9.2 Archival Capabilities to Handle Machine-Readable Records

The second requisite of a MRR program is for the archival agency to have the capability to appraise, accession, process, preserve, and provide reference services for MRR. This requires new skills by the archives staff,

and the capacity to store and disseminate MRR. Archivists must obtain a rudimentary knowledge of computing, gain familiarity with the terminology used by data processing personnel, learn basic technical procedures for accessioning and preserving the records, and understand the information system that MRR can support.

9.3 Access to Outside Technical Skills and Resources

Archivists must identify available outside resources and expertise. These resources include computing facilities needed to accession, process, maintain, and perhaps store the records. Likely sources of such facilities are computer service centers for state agencies, university data library and computing facilities, and private service bureaus. Criteria for selecting a service bureau include compatability of the hardware and software with that used by state agencies, the availability of software packages for manipulating the data, the quality and availability of the computer center's technical support staff, and costs.

The archives staff must develop working relationships with technical experts. Archivists will find it necessary to seek their advice for resolving specific data structure and technical problems and for guidance in shaping policies that will be affected by technological change. In some instances, the archives will be required to consult free-lance programmers or systems analysts.

Archivists must also develop working relationships with university and agency researchers who know the methodology and research trends in specialized fields. Many MR files will have secondary applications which differ from their primary purposes. The advice and guidance of such researchers in appraising these files for their potential use can improve the quality of appraisal decisions.

9.4 Legislative and Administrative Guidelines to Govern Access

There is a commitment on the part of the archivist to make all records as freely accessible as possible without violating laws or without compromising an individual's right to privacy. MRR make it technically possible to carry out this archival commitment. However, there is little in the statutes to guide the records creator, the archivist, and the researcher in the use of MRR which contain personal identifiers.

Because archival records have little value if they cannot be used, a final requisite of an effective MRR program is a legislative recognition of the legitimate role of the archives in providing access to confidential MRR. Such legislation must delegate to the archives the responsibility for making MRR available for scholarly research while also assuring that appropriate safeguards to protect individual rights of privacy are created and maintained.

Ideally, such legislation should have some sort of "sunset" clause which would automatically lift restrictions on records after a specified period. Or restrictions could be lifted after a specified period if reviewed and approved by a public records board or an open records board.

10. Elements of a Machine-Readable Records Program

Each state archives will have a different capacity to develop and implement a MRR program. Factors that will influence this capacity include the nature of existing programs for textual records, the availability of skills and resources within the archives and from related outside organizations, and the amount of cooperation that can be elicited from central records management staffs and agency personnel. Nevertheless, an archival program for state MRR must include three components: pre-archival control, archival preservation, and research use and access. In general, the three components are incremental and can be implemented in stages. However, there are many occasions when they overlap.

10.1 Pre-archival Control

The nature of MRR requires that different strategies for their control be employed. MRR must be identified and controlled at an early stage in their life cycle. The pre-archival control activities must consist of identification, preventing unauthorized destruction, archival review, and agency preservation. One of the first functions is identifying the records. During the initial phases of establishing a MRR program, archival agencies will probably be required to take responsibility for initiating and conducting surveys. Such surveys might concentrate on systematic assessments of all MRR in entire agencies, on specific subject areas, on systems that are known to exist, or on data files that are in particular danger of destruction or deterioration. Whenever surveys of textual records are being conducted, MRR should be included.

These inventories are only a short term solution. The key to an on-going program is agency personnel and records managers who identify and schedule MRR as they are created. These surveying activities should be integrated into scheduling activities for non-MRR. During the initial phases, the archives staff must assume responsibility for training agency records management and data processing personnel about the legal status and importance of public records in MR form. Records managers will need training in techniques to identify, inventory, describe and schedule MRR. Data processing personnel must learn about the importance of long-term MRR maintenance and preservation. This training can be provided through general seminars and workshops and through individualized instruction. Records managers should be asked to assist the archives staff in conducting surveys, so that they can learn the survey techniques.

A second objective must be to prevent unauthorized destruction of the records. Until agency personnel become accustomed to scheduling MRR, it may be necessary to impose a blanket stop order on MRR destruction until inventories can be made and archival appraisals completed. Such an order, which would surely be unpopular in computer centers and very difficult to enforce, should be a last resort. Other measures include more informal means: (1) personal contacts with computer center managers; newsletter articles and circulars; and presentations before groups of program managers and data processing personnel, all of which is aimed at education; and (2) negotiating the right to informally review computer center "scratch" requests. These are stop gap measures until more systematic means of review are developed.

10.2 Archival Review

Identification, scheduling and preventing destruction makes it possible for the archival agency to appraise the records. Ideally, information upon which the archivist's appraisal decision is based will be a records schedule. The ideal records schedule will consist of several entries, one for each series of records, which describe the related parts of a records-keeping system.[6] Each entry consists of three parts: a concise description of the series and its relationship to other records; a retention period (the period of time the agency needs the records for functional, administrative, analytical, legal or fiscal purposes); and a disposition request (destroy or transfer to archives). The retention period is determined by the agency (and is reviewed by legal and fiscal authorities). But the final disposition is determined by the archival agency based on its appraisal of the records.

Because many MRR contain micro-level data, archivists must reevaluate a traditional appraisal principle: That aggregated records or summaries should be accessioned in lieu of the micro-level records. Both the bulk and difficulty of using micro-level records in their hard-copy form would dictate that they not ordinarily be accessioned. However, micro-level MRR do not present a significant space problem and their form makes sophisticated detailed analysis possible, which is rarely practicable for paper records.

Appraisal of MRR must also take account of technical considerations:

(1) Is documentation available or can it be easily reconstructed? If not, the records should not be scheduled for transfer to the archives.
(2) Are the records in a physical form that makes their transfer possible? (Are they recorded on an outmoded storage medium? Has the

---

[6] i.e., the source document; data entry forms; input transaction files; error listings; edit sheets; proof sheets or "dumps" of the input; the MR master file; MR subsets of the master file; interim reports; and final reports.

storage medium deteriorated? Are the records kept in a data base system which would make it difficult to select non-current records?)

(3) What will be the costs of accessioning, processing, and storing these records?

Answers to these question will help the archivists arrive at a preliminary decision about whether or not to accession a MRR series. When the time to actually accession the records arrives additional factors may be discovered: the informational quality of the records might not be as rich as supposed; the documentation, although available, might be inadequate; or the costs to accession and process may have been underestimated.

Another pre-archival control activity is agency preservation. The archival agency, with agency records management, computer centers, and records center personnel, should establish policies and procedures for storing MR files under proper environmental and security conditions and for maintaining their associated documentation. Either agencies take responsibility for long-term preservation of MRR and documentation, or the archives negotiates early transfer to the archives of a security copy.

10.3 Archival Preservation and Management of Machine-Readable Records

Archival preservation will require policies and procedures for preserving and maintaining the MRR and training archival staffs so that they have the technical skills to handle MRR. Although MRR are stored on a variety of media in the agency, the archival storage medium will be magnetic tape. Systematic maintenance is essential to insure magnetic tape's preservation. To prevent the loss of data due to damaged, unreadable, or lost magnetic tapes, both a master and security back-up copy of any MRR file should be generated on new magnetic tape. If the information on one copy becomes inaccessible, it should be possible to recover the data from the other. The master and security copies should each be stored at separate physical locations.

A variety of procedures should be implemented to maintain relatively constant tension levels for long periods of storage. Tapes should be cleaned and rewound on an annual or biennial basis. Periodic recopying of data to new tape is crucial for preserving the MR data. The need for frequent recopying of magnetic tapes is in large part dependent upon environmental storage conditions. As with other archival media, a controlled environment is essential; it should be relatively dust free, protected from high intensity magnetic fields, and constant in temperature and humidity levels. The magnetic tapes should be stored in an upright position in plastic containers.

In addition to the potential physical deterioration of the storage medium, technical advances may result in the obsolescence of the storage format and/or medium. Given the potential costs involved, it is doubtful that maintaining obsolete hardware and software is a feasible alternative. A more viable alternative is conversion from one storage format and/or medium to the current standard format and/or medium.

As a result of dynamic technological changes during the past few years, a variety of alternative media for storing MRR are being developed, which are based on other technologies, such as optic or electron-beam technologies. Archivists must keep informed about these changes, although it may be some time before new devices with long-term archival properties are available.

The complexity of data bases and contents require access to a sophisticated set of methodological and technical tools. The creation of an archival record will require greater agency cooperation than was heretofore necessary. Assessing the utility of data bases for future scholarly applications will also require access to expert users, to advise archivists on the potential uses of the data. Assessments of the contents of the records and of the structure of the data files should be made in concert with advisory committees to the archives. (This process has been successful for the creation of useful federal administrative records and public use files.)

Data security is another issue the archives must be concerned with. Some MRR transferred to the archives may contain confidential information. A systematic procedure to restrict access to such data must be devised and implemented. While archives should have control over access to any tapes within their holdings, archivists will have to depend on security and access protection systems available at computer installations to prevent unauthorized use of the data. Such a security system might include the following:

(1) Designation of all MRR by type of access.
(2) Control over the tape library by a single person responsible for security of all materials therein.
(3) Transfer of all materials to be read at the computer under appropriate security and control.
(4) Pre-coded passwords to assure that faulty mounting of tapes does not result in inadvertent disclosure of contents to unauthorized persons.

Unrestricted data must be handled with the same security as restricted files to assure that error does not cause accidental destruction of tapes or use by unauthorized persons.

The importance of documentation in order to understand, retrieve, and manipulate MRR, has already been noted. Documentation, either supplied by the agency or prepared by the archives, should consist of the following:

(1) References to relevant statutes and program guidelines which authorize the creation of the MRR.
(2) The source documents which provide the basis for data collection, entry, and processing (to assist the archivist and researcher in determining to what extent the MRR reflect the original data gathering activity).
(3) Information on sampling and data collection procedures (to assist in making appraisal and retention decisions and for making an informed judgment about the applicability of the MRR to a particular research project).

(4) Information on the processing activities by which the MR file was created, updated, corrected, and changed (to determine the quality of the editing and checking procedures, changes in definitions of the data elements, observations, levels of aggregation, and the reliability of the information).
(5) Information on the physical organization of the MRR; including how the records are structured, in what form they are written, and how they must be transferred in order to be accessed and their contents retrieved.
(6) Information on the organization of the data elements, commonly referred to as a "codebook" or record layout (to provide an understanding of the relationship between the source documents and the MRR, to locate each data element in the file, and to provide an indication of the quality of the data file. In addition, the codebook describes the coding structure for each data element, which is essential in order to manipulate the data elements for statistical analysis and interpret quantitative and categorical data.)
(7) Printouts of several records from the file (to provide a picture of the records and to provide the archivist with additional information on the MRR--particularly when information on their contents is not provided by a codebook).
(8) Reports or products (or citations to such reports) generated by the MRR (to assist in making appraisal and retention decisions, because so often information about MRR is inadequately described elsewhere).
(9) Information on rules governing access to MRR (to help make decisions about transferring records to the archives and about the conditions that govern access by users).
(10) Information on the nature of the computer and software environment in which the MRR are located (to determine how and whether the MRR can be transferred to and preserved by the archives and in what form they can be accessed by users).

A MRR program will incur a different set of costs for the archival agency. A routine preservation program is less costly and more effective than an emergency program designed to salvage information recoverable from deteriorated media. The costs of a maintenance program include capital equipment, computer time, and personnel. If MRR are already recorded in standard transfer formats and the essential technical documentation required to access and retrieve the data are readily available, the cost of maintaining and preserving the data will be small.

The cost for personnel and computer time when MRR cannot be easily transferred from agency custody to the archives can be high. Specialized staff will be required to produce archival MRR. It may be necessary, if the archives does not have specialized staff, to purchase the services of agency analysts and data processors or to use the staff at a nearby computing center or computer service bureau. In any case, the budget for a program of MRR archival preservation will require allocation of funds for technical staff, computational facilities, and capital equipment for storage.

10.4 Access and Use

Providing access to the archival MRR will require developing relationships with technical support and user communities.

The archives must provide physical access to the MRR. This can be done via a computer center or service bureau, or a social science data archives. The archives may choose to store a user copy of the MRR at a computing center. After a user decides which MRR to use, the archives makes a copy of the data file available to the user at the computing center. This strategy is efficient both from the archival and the user's perspective, especially if the computing center has available a wide range of data management and statistical software and a technical support staff who can assist the user in carrying out the research project. This strategy relieves the archival repository of the burden of providing technical user services. The user bears the cost of access, retrieval, and manipulation of the data, and perhaps, depending on the policies of the computer center, the cost of technical support. The only cost to the archives is the storage of the tapes.

Social science data archives offer another alternative to providing user services. They are typically located at universities or colleges, and are experienced in providing access tools and technical support for MRR users. Archival MRR can be deposited at a social science data archives, which acts as the disseminating agent for the public records archives. This sort of arrangement calls for establishing policies and procedures to share the archival responsibilities for appraisal, accessioning, processing, description, and dissemination.

In some cases, access and use of MRR may depend on their contents. MRR may contain confidential records, for which access depends on state statutes and the originating agency's policies with regard to these records. In cases where the agency has delegated access responsibilities to the archives, the archives will need to establish policies and procedures for use of these MRR (see data security, p. 46). Among the policies will be determining what information can be released to users. The archival agency will need to obtain guidance from experts in statistics and other disciplines in order to determine what strategies are required for masking the identity of the individual records, so that sensitive information is not released. Among the policies should be those which require the user not to reveal individual identities of any of the cases and to bear the responsibility for assuring that no harm may come to the individuals through inadvertent disclosure.

Increasingly, MRR are created in complex and dynamic environments that require new and different user communities to provide expertise on creating useful MRR for future research and scholarly activities. Because automated information management systems present a host of intellectual and technical problems, the archival agency will want to work closely with state agency and academic users to determine how archival MRR can be produced from these new environments. While it is impossible to predict with any certainty what research use will be in the next decades, guidance from the research community

will be invaluable to the archivist in deciding what data to retain from sophisticated information management systems.

11. Concluding Remarks

The effectiveness of archivists in this era of rapid technological change depends on their ability to alter past behavior and to fashion strategies to cope with both the opportunities and the problems created by change. Technology and increased record-keeping by government motivates archivists to reexamine many basic assumptions about archival theory and practice. As the project findings have demonstrated, many assumptions must be revised if a record of governmental activities is to be maintained.

Our project demonstrated that cooperation must become a central archival strategy for the preservation of MRR. Interinstitutional cooperation is a requirement, fostered by the complex environment and technology of which the archives is a part. Programs for conserving the archival MRR will require increased cooperation between governmental agencies and the archives, between the archives and other organizations which are specialists in this type of record, and between the archives and the user community.

Preserving archival MRR will also require that the archives engage in new planning strategies for identifying and analyzing records needs, delineating objectives, devising and testing new approaches, and evaluating its achievements. The archival profession must educate itself in the preservation and use of machine-readable records.